CS4414. Supercomputing Allocation Panel Mock Review 2

You are approached by a team of molecular biologists from Harvard Med., school. The team may be on a verge of a revolutionary discovery: cure for HIV. Their key idea looks very promising (4 papers in "Science" and 3 in "Nature" just this last year). The gist of their proposal is a drug that is supposed to "zap" the viral genes even once they become part of the infected person's genome. The reason HIV is such a nasty disease is that the virus infects the victim's genome itself, unlike, say, flu.

The drug efficacy evaluation protocol consists of several stages, but the total can not be more than 24hrs long. In fact, because the "biological" parts take long, the computational part must be done within 1 hr. Here is what needs to be done: the whole genome of a patient about 10^{10} letters of the DNA code (obviously, in digital form at this stage, assume a string of "0" and "1"), needs to be screened against the known HIV genome, about 10^4 letters, to find how many matches are there in the patient genome before and after the administration of the drug. If the number of matches decreases substantially, say by at least a factor of 10, it is a success! The decrease does not have to be computed accurately, all the team needs to know if it is large or not. The team is going to use the best available algorithm for the task, which can process 10^5 letters per hour. Unfortunately, no parallel implementation of it is available, and there is no time to develop one of high quality.

Your task is to advice the team and make sure the new therapy becomes a reality.